ERGDB: Estrogen Responsive Genes Database

Suisheng Tang, Hao Han and Vladimir B. Bajic*

Knowledge Extraction Laboratory Institute for Infocomm Research, Singapore

ABSTRACT

ERGDB is an integrated knowledge database dedicated to genes responsive to estrogen. Genes included in ERGDB are those whose expression levels are experimentally proven to be either up-regulated or down-regulated by estrogen. Genes included are identified based on publications from the PubMed database and each record has been manually examined, evaluated and selected for inclusion by biologists. ERGDB aims to be a unified gateway to store, search, retrieve and update information about estrogen responsive genes. Each record contains links to relevant databases, such as GenBank, LocusLink, Refseq, PubMed and ATCC. The unique feature of ERGDB is that it contains information on the dependence of gene reactions on experimental conditions. In addition to basic information about the genes, information for each record includes gene functional description, experimental methods used, tissue or cell type, gene reaction, estrogen exposure time and the summary of putative estrogen response elements if the gene's promoter sequence was available. Through a web interface at http://sdmc.i2r.a-star.edu.sg/ergdb/cgi-bin/explore.pl users can either browse or query ERGDB. Access is free for academic and non-profit users.

INTRODUCTION

Estrogen is recognized as a key regulator of growth, differentiation and metabolism in animals. It targets multiple organs such as brain, heart, bone, breast, uterus and prostate. In humans, estrogen is associated with many diseases, including osteoporosis, obesity, breast cancer, arteriosclerosis and Alzheimer's disease (1,2). Generally, the effects of estrogen are mediated through its binding to estrogen receptors (ERs) that function as transcription factors regulating the expression of estrogen responsive genes (3,4). Irrespective of the induction pathways used, the final effect of estrogen is exerted through the activities of the estrogen target genes. For this reason the knowledge of estrogen responsive genes becomes essential for clinical and basic research for issues and diseases related to imbalances of estrogen, gene networks controlled by estrogen, transcription regulation of target genes, diagnostic targets and drug targets for implied diseases. Identification and

understanding of the transcriptional mechanisms of genes whose transcription is regulated by estrogen is, thus, of fundamental importance.

The transcription of many genes is modulated by estrogen. Microarray studies (5,6) showed that estrogen affects hundreds of genes in diverse cell lines and tissues. The gene expression profiles change in various developmental stages and disease states (7,8). In fact, a gene's responses to estrogen are dependent on many factors, including the available subtype of ER, the co-regulators, the estrogen exposure time and the amount of estrogen (9,10). Experimental information is thus essential for any further analysis. It is the foundation for the study of gene expression. Analyzing such information we may find clues related to estrogen involvement in the mechanisms of cell proliferation, differentiation, cancer progression and treatment, etc.

Information about estrogen responsive genes has been accumulated rapidly due to the efforts of the scientific community and the improvement of high-throughput experimental technologies. However, the rapidly increased quantity of data poses problems for individual researchers. For example, a simple query in PubMed using 'estrogen AND gene' yields over 9500 records. Facing a huge number of records, a researcher has difficulty in selecting relevant records. It is virtually impossible to quickly (and accurately) get an overview of genes regulated by estrogen. Another inconvenience for the researcher is scattered sources of data. Information on gene biological functions, gene sequences and literature reports are spread across several databases. One of the great problems is the management and integration of that high volume of information. We believe that an integrated database equipped with useful analysis tools and with online access contributes to the solution. Although there are several hormone-related databases, to the best of our knowledge currently there are no databases focused on hormone responsive genes. For example, NUREBASE (11) (http://www.ens-lyon.fr/LBMC/laudet/nurebase.html) contains information about nuclear hormone receptors, while NRMD (12) is a database focused on mutation of nuclear receptors. Neither of these focuses on hormone inducible genes. To complement the existing databases and to contribute more efficiently to the functional genomics field, we initiated a project on Genes Responsive to Nuclear Hormones and developed the Estrogen Responsive Genes Database (ERGDB) as the first milestone of the project. ERGDB aims to be a unified gateway to store, search, retrieve and update information about estrogen responsive genes in order to support research in steroid hormones, gene networks and functional genomics in general.

---

*To whom correspondence should be addressed. Tel: +65 6874 8800; Fax: +65 6774 8056; Email: bajicv@i2r.a-star.edu.sg

Table 1. Links mentioned in ERGDB

Name	URL
PubMed	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed
LocusLink	http://www.ncbi.nlm.nih.gov/LocusLink/
RefSeq	http://www.ncbi.nih.gov/RefSeq/
Genbank Genome Resource	http://www.ncbi.nlm.nih.gov/Genomes/
Dragon ERE Finder 2.0	http://sdmc.i2r.a-star.edu.sg/ERE-V2/index
ATCC	http://www.atcc.org/SearchCatalogs/CellBiology.cfm
NUREBASE	http://www.ens-lyon.fr/LBMC/laudet/nurebase.html

ERGDB is an integrated knowledge database where every record has been selected and evaluated manually by biologists. ERGDB provides basic gene sequence information and description, but it highlights experimental conditions, such as laboratory methods, animal models, tissue or cell type, estrogen exposure times and gene reactions. Experimental information is important for any further research because it provides the foundation on which to compare gene reactions as well as to judge the accuracy of the results. Live links to relevant sources of data and information are provided. We integrated with ERGDB a powerful tool for locating estrogen response elements (EREs) in the promoter regions of respective genes (13). We also provide the details of the putative EREs in the genes included in ERGDB. This information may prove helpful for functional genomics research. ERGDB is freely accessible at http://sdmc.i2r.a-star.edu.sg/ergdb/cgi-bin/explore.pl for academic and non-profit users.

DATABASE OVERVIEW

The goal of ERGDB is to provide an integrated knowledge database that allows scientists, as well as clinicians, to get a global overview of the known information concerning genes regulated by estrogen. Records in ERGDB are selected, evaluated, quality-controlled and completed by biologists. This makes it different from databases that are populated in an automated fashion. We emphasize that gene expression depends on experimental conditions. This information is mostly not available in the existing databases. Each record in ERGDB provides details of experimental conditions collected from the literature, such as tissue or cell type, estrogen treatment time-point, gene reactions, laboratory methods used. Users of ERGDB can obtain answers to questions such as:

(i) is the gene I am interested in regulated by estrogen (or which genes are regulated by estrogen)?;

(ii) how many and which genes are regulated by estrogen in a specific tissue or cell line?;

(iii) how many and which genes are affected by estrogen at a particular time-point? (This last question cannot be answered by a single query, but information can be collected from individual responses.)

According to the experimental conditions users can compare gene reactions and judge the accuracy of results. Conveniently, relevant links and an analysis tool for locating putative ERE patterns are also available in the database.

Estrogen responsive genes are defined here as genes whose expression has been significantly either up-regulated or down-regulated by estrogen. To be included in the ERGDB a gene must satisfy the following four conditions:

(i) Gene expression has to be experimentally proved to be either up-regulated or down-regulated by estrogen. Data from any animal models and cell lines are included.

(ii) The gene must exist naturally. Artificially constructed genes are not included.

(iii) The change in expression levels of the gene must be caused by estrogen only. Genes affected by a combination of estrogen and other agents, such as tamoxifen, are not included.

(iv) The gene expression level has to be affected significantly. For microarray experiments, a gene is included only when its expression change is $\geqslant 1.5$-fold. Results from other methods, such as RT–PCR, *in situ* hybridization and northern blot analyses, have to be statistically significant.

Protein concentration variations or enzyme activity changes due to the estrogen treatment are not included since their changes may be affected by many factors rather than only being regulated at gene expression level.

Information in ERGDB is organized into three parts:

(i) General information, which includes organism, gene name, gene synonyms, gene description, chromosome, strand, contig location and cross-reference links. This part of the information helps user to get a basic gene information and serves to support further sequence analysis. It also contains links to the Locus Link, Unigene and Refseq databases.

(ii) Experimental information, which includes experimental method, model, regulation effect of gene expression, cell line/tissue type, estrogen exposure time and the original reference. This part of the information details experimental conditions and can support building gene interaction networks or hormone reaction cascades. Information about cell line/tissue type is linked to the American Type Culture Collection (ATCC). Most conveniently, the abstract of the original publication is accessible via a PubMed link.

(iii) Putative ERE pattern information, which contains data based on the analysis of a gene's promoter region by the Dragon ERE Finder v.2.0 system (13). This information is useful for gene regulation studies since ERE is a specific fragment in the promoter regions of estrogen responsive genes to which ERs bind. Although ERs can bind to other sites in the promoters, the presence of a strong ERE pattern is a good indicator that the gene may be regulated by estrogen. We provide detailed information about putative ERE patterns, such as the ERE location in the strand, its position in the contig, its composition and whether it is a new or known ERE pattern. However, ERE pattern information is provided only if we can accurately determine the start of the gene. To facilitate further analysis of the ERE patterns and surrounding regions, we have provided the Dragon ERE Finder tool in the database.
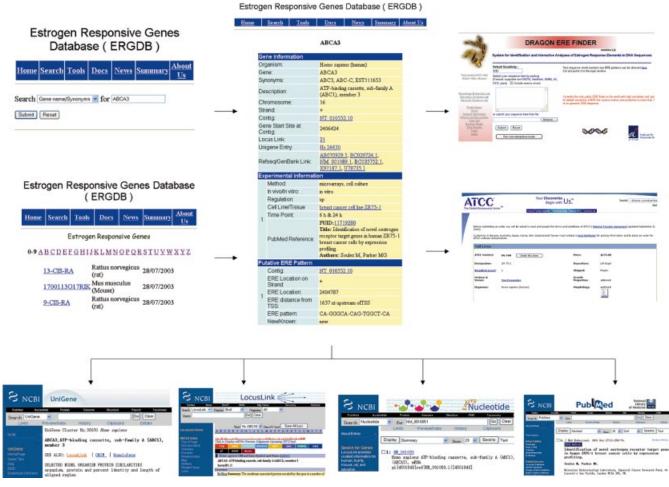
Figure 1. Example of an ERGDB record and its relationship to other links.

HOW TO USE ERGDB

To search for target genes, users can type in, on the search page, one of the following gene identifiers: LocusLink (14) ID or GenBank (15) accession number, or full gene name or part of it. If the gene is contained in ERGDB, a table of information will be returned as in the example in Figure 1. If the gene information is not included in ERGDB, a statement that the item is not found will be issued. To find all the genes regulated in a target cell line or tissue included in ERGDB, users should select 'Cell line/Tissue' for searching and type in the cell line/tissue name. ERGDB will return a list of genes regulated by estrogen in that specific cell line or tissue. To view a list of all genes collected in ERGDB, users can visit the home page and click the 'List of estrogen responsive genes' at the bottom of the page. Gene names are shown in alphabetical order. To find out more details about ERGDB and its possibilities please consult the ERGDB website and relevant FAQ page.

METHOD AND IMPLEMENTATION

ERGDB stores information about estrogen responsive genes. Since each record is selected based on published experimental evidence, original scientific publications have been the best data source. The PubMed database has been the starting point in our data collection. PubMed covers nearly 4500 journals and contains more than 12 million abstracts, mainly in life sciences. We screened PubMed using queries with keywords. A typical query consisted of two parts: hormone name ('estrogen' or 'estradiol' or 'oestrogen' or 'oestradiol') and gene reaction status ('up regulat*' or 'up-regulat*' or 'down regulat*' or 'down-regulat*'). Publications that were matched by these queries were then examined by biologists (either abstracts or whenever possible full articles). Since ERGDB contains detailed information about the experimental conditions under which gene regulation by estrogen has been proved, we had to resort to manual evaluation of the published data and this is one of the characteristics of ERGDB. Although we could use text mining techniques, these were not applied for information extraction in our case since they did not show sufficient accuracy. Briefly, four steps have been carried out after the initial publication screening. (i) Information about the experiment, such as animal model, cell line or tissue type, estrogen exposure time, gene expression results, experimental methods, is carefully evaluated and recorded. (ii) Once the experimental information is confirmed, gene information is collected from the relevant public databases. (iii) A search for ERE patterns was performed if the gene sequence and accurate

gene start information were available. (iv) All data were integrated and organized in such a manner that users can efficiently query and browse information.

CURRENT STATUS AND FUTURE DEVELOPMENTS

We have manually examined more than 5000 PubMed records, dated from 1995 to March 2003, for genes whose expression levels are significantly changed by estrogen treatment in various experiments. The current release (v.1.0) of ERGDB contains information about 797 different genes, 236 different experiments and related results, and references to 146 original publications. Records for human genes are generally with the complete gene information. Information for genes from other species may be lacking some of the information, such as strand, etc. since this information was not complete in public databases. As a result the ERE patterns may not be available. As the sequence data for different species becomes more complete, so will the updated ERGDB content. ERGDB is designed to adapt to and reflect the most current information about estrogen responsive genes. It will continue to grow in both content and functionality and will be updated every month to include any new data from literature.

REFERENCES

1. Jordan,V.C., Gapstur,S. and Morrow,M. (2001) Selective estrogen receptor modulation and reduction in risk of breast cancer, osteoporosis, and coronary heart disease. *J. Natl Cancer Inst.*, **93**, 1449–1457.
2. LeBlanc,A. (2002) Estrogen and Alzheimer's disease. *Curr. Opin. Investig. Drugs*, **3**, 768–773.
3. Nilsson,S. and Gustafsson,J.A. (2000) Estrogen receptor transcription and transactivation: Basic aspects of estrogen action. *Breast Cancer Res.*, **2**, 360–366.
4. Nilsson,S., Makela,S., Treuter,E., Tujague,M., Thomsen,J., Andersson,G., Enmark,E., Pettersson,K., Warner,M. and Gustafsson,J.A. (2001) Mechanisms of estrogen action. *Physiol. Rev.*, **81**, 1535–1565.
5. Inoue,A., Yoshida,N., Omoto,Y., Oguchi,S., Yamori,T., Kiyama,R. and Hayashi,S. (2002) Development of cDNA microarray for expression profiling of estrogen-responsive genes. *J. Mol. Endocrinol.*, **29**, 175–192.
6. Seth,P., Porter,D., Lahti-Domenici,J., Geng,Y., Richardson,A. and Polyak,K. (2002) Cellular and molecular targets of estrogen in normal human breast tissue. *Cancer Res.*, **62**, 540–544.
7. Reese,J., Das,S.K., Paria,B.C., Lim,H., Song,H., Matsumoto,H., Knudtson, K.L., DuBois,R.N. and Dey,S.K. (2001) Global gene expression analysis to identify molecular markers of uterine receptivity and embryo implantation. *J. Biol. Chem.*, **276**, 44137–44145.
8. Soulez,M. and Parker,M.G. (2001) Identification of novel oestrogen receptor target genes in human ZR75-1 breast cancer cells by expression profiling. *J. Mol. Endocrinol.*, **27**, 259–274.
9. McDonnell,D.P. and Norris,J.D. (2002) Connections and regulation of the human estrogen receptor. *Science*, **296**, 1642–1644.
10. Hall,J.M., Couse,J.F. and Korach,K.S. (2001) The multifaceted mechanisms of estradiol and estrogen receptor signaling. *J. Biol. Chem.*, **276**, 36869–36872.
11. Duarte,J., Perriere,G., Laudet,V. and Robinson-Rechavi,M. (2002) NUREBASE: database of nuclear hormone receptors. *Nucleic Acids Res.*, **30**, 364–368.
12. Van Durme,J.J., Bettler,E., Folkertsma,S., Horn,F. and Vriend,G. (2003) NRMD: Nuclear Receptor Mutation Database. *Nucleic Acids Res.*, **31**, 331–333.
13. Bajic,V.B., Tan,S.L., Chong,A., Tang,S., Strom,A., Gustafsson,J.A., Lin,C.Y. and Liu,E.T. (2003) Dragon ERE Finder version 2: a tool for accurate detection and analysis of estrogen response elements in vertebrate genomes. *Nucleic Acids Res.*, **31**, 3605–3607.
14. Pruitt,K.D. and Maglott,D.R. (2001) RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.*, **29**, 137–140.
15. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2003) GenBank. *Nucleic Acids Res.*, **31**, 23–27.